

Chenhui Shao
Department of Mechanical
Science and Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801
e-mail: chshao@illinois.edu

Jie Ren
Department of Industrial and
Manufacturing Engineering,
Florida State University,
Tallahassee, FL 32310
e-mail: jr14r@my.fsu.edu

Hui Wang¹
Department of Industrial and
Manufacturing Engineering,
Florida State University,
Tallahassee, FL 32310
e-mail: hwang10@fsu.edu

Jionghua (Judy) Jin
Department of Industrial and
Operations Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jhjin@umich.edu

S. Jack Hu
Department of Mechanical Engineering,
University of Michigan,
Ann Arbor, MI 48109;
Department of Industrial and
Operations Engineering,
University of Michigan,
Ann Arbor, MI 48109
e-mail: jackhu@umich.edu

Improving Machined Surface Shape Prediction by Integrating Multi-Task Learning With Cutting Force Variation Modeling

The shapes of machined surfaces play a critical role affecting powertrain performance, and therefore, it is necessary to characterize the shapes with high resolution. State-of-the-art approaches for surface shape characterization are mostly data-driven by interpolating and extrapolating the spatial data but its precision is limited by the density of measurements. This paper explores the new opportunity of improving surface shape prediction through considering the similarity of multiple similar manufacturing processes. It is a common scenario when the process of interest lacks sufficient data whereas rich data could be available from other similar-but-not-identical processes. It is reasonable to transfer the insights gained from other relevant processes into the surface shape prediction. This paper develops an engineering-guided multitask learning (EG-MTL) surface model by fusing surface cutting physics in engineering processes and the spatial data from a number of similar-but-not-identical processes. An iterative multitask Gaussian process learning algorithm is developed to learn the model parameters. Compared with the conventional multitask learning, the proposed method has the advantages in incorporating the insights on cutting force variation during machining and is potentially able to improve the prediction performance given limited measurement data. The methodology is demonstrated based on the data from real-world machining processes in an engine plant. [DOI: 10.1115/1.4034592]

Keywords: multitask learning, Gaussian process, data fusion, surface variation, surface modeling

1 Introduction

The control of surface shape variation on a machined surface is of great importance in automotive powertrain manufacturing because such variation has a significant influence on the sealing performance and causes distortion during surface assembly, such as those between engine heads and blocks or upper and lower transmission valve bodies. Characterization of the surface shape with high precision has become critical in controlling surface variations to ensure the functional performance of surface assembly.

Various high-resolution surface measurement systems have been developed, including profilometers and laser holographic interferometers. However, the acquisition of high-resolution measurements using those systems is time-consuming and expensive. Moreover, such measurements may not be robust in response to environmental disturbance. For example, the measurement accuracy based on a laser holographic interferometer system can be affected by the heat dissipated from a machined part, measurement table vibration, and surface contamination, thus resulting in data loss and measurement errors, as shown in Fig. 1. More recent advancement in surface sensing technology is reviewed by Wilton et al. [1]. However, fast and high-resolution surface

metrology systems have not been widely adopted in quality control for powertrain production.

Various approaches have been developed to make fine-resolution evaluation of the entire surface shape based on low or multiresolution surface measurements. These approaches include interpolating and extrapolating surface shapes from measured points through least squares [2], B-spline methods [3,4], and grid fit through triangulation [5]. In spatial statistics, the method is to estimate the surface shape by considering spatial correlations among sampled points. Such correlations reflect the spatial similarities between data in the vicinity on the part surface, and have

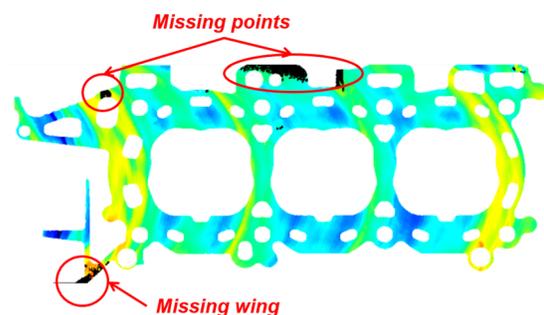


Fig. 1 Data loss in high-resolution measurement

¹Corresponding author.

Manuscript received April 23, 2016; final manuscript received August 5, 2016; published online September 29, 2016. Assoc. Editor: Radu Pavel.

been extensively utilized to interpolate and extrapolate surface data for form error estimation in machining processes [6–8], modeling of spatial wafer variations [9,10], and geographic understanding in remote sensing applications [11,12]. One limitation with these data-driven approaches is that the precision of the surface predictions is constrained by the density of measurements. To deal with the limitation, process knowledge can be integrated with surface measurements to cost-effectively model surface shapes [13,14]. The prediction improvement was achieved by fusing multiresolution measurements with cutting force dynamics during surface manufacturing. In summary, state-of-the-art surface modeling techniques mostly focus on the interpolation/extrapolation of the data and process information from *one single manufacturing process*. Additionally, most of the data-driven methods lack engineering insights.

This paper will explore the new opportunities of improving surface shape modeling by considering the commonality between multiple manufacturing processes. The idea is motivated by a common manufacturing scenario when the process of interest lacks sufficient data, whereas rich data could be available from other similar-but-not-identical processes. For instance, an auto-maker plans to build a new engine plant where there is a lack of historical data to establish a baseline for surface quality control. In the meantime, other engine plants have been running for decades, and therefore, the relevant historical surface data are extensively available. Though the face milling processes in the existing plants were developed for different types of engines, the cutting mechanism for the same materials could be very similar, and thus, those process data can partially contribute to the prediction of the surface shapes in the new plant (Fig. 2(a)). A similar scenario applies to a parallel machining station where multiple machines produce the same type of parts in parallel (Fig. 2(b)). Surface data from the parallel machines can be utilized to jointly learn the surface shape models for all machines.

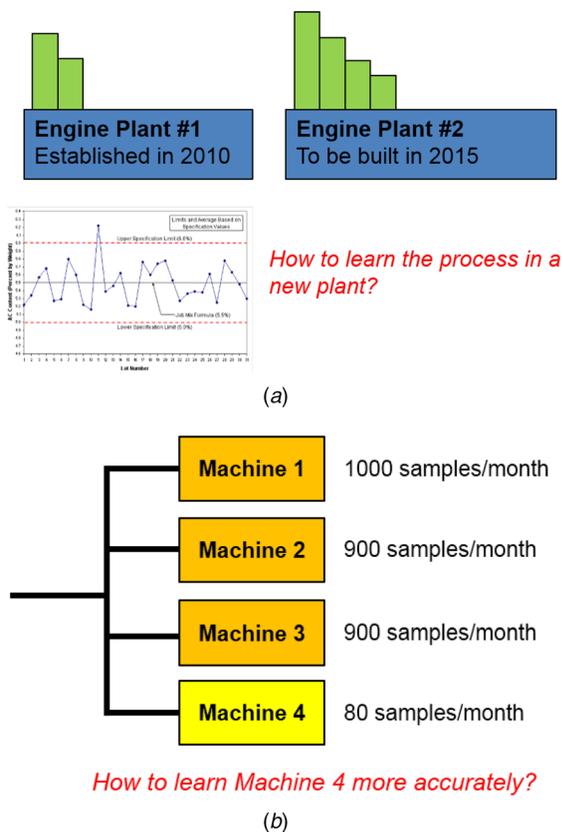


Fig. 2 Potential knowledge transfer in manufacturing applications: (a) new plant and (b) low production station

These examples motivate us to improve the surface variation prediction by transferring the knowledge learned from data-rich processes to a similar process. Multitask learning has emerged as a solution to knowledge-transfer problems [15]. It can be particularly useful in the situation when a limited amount of data is available in the target task, while data from other related tasks are readily available. Here, a task refers to the learning/estimation of model parameters using training data. It is expected that by learning these tasks simultaneously, superior prediction performance can be achieved over the “no transfer” case (i.e., when each task is learned in isolation). Figure 3 summarizes the difference between multitask learning and traditional single task learning. Multitask learning has been employed to Gaussian process prediction, neural network, Dirichlet processes, and support vector machine [15,16]. The key problems of multitask learning include the modeling of the similarity among multiple processes and the learning algorithms for model parameters. As for Gaussian processes, various similarity models such as block diagonal or not block diagonal covariance matrix and learning algorithms such as expectation-maximization algorithm are proposed [17–19]. Therefore, it can be anticipated that through multitask learning, the quality performance of the manufacturing process of interest or the new plant can be predicted by utilizing the observations from different but related manufacturing processes or plants.

It should be noted that the current multitask learning for Gaussian processes is data-driven and does not incorporate the engineering insight into the model prediction. For example, our prior research has uncovered a positive correlation between surface height and effective material removal rate [20,21] through cutting force dynamics and spatial data modeling. Incorporation of the cutting force variation induced surface modeling can help to improve surface shape prediction. The paper aims to improve surface shape prediction by fusing cutting force variation modeling with surface measurement data from a number of similar machining processes via iterative multitask Gaussian process learning. We first establish a surface model consisting of a global trend that is induced by cutting force dynamics and a local variation term that can be characterized by a zero-mean Gaussian process. The global trend captures the overall shape variation across the

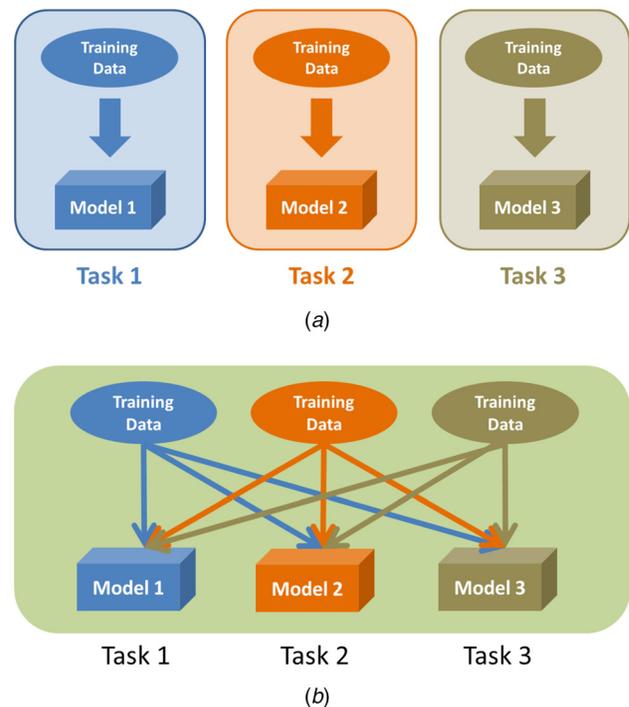


Fig. 3 Difference between single task learning and multitask learning: (a) single task learning and (b) multitask learning

surface, and can be well explained by the process physics [13]. An iterative multitask Gaussian process learning algorithm is then developed to learn the parameters from the global trend and the local variation. The algorithm will be demonstrated based on engine head deck face machining processes.

The remainder of this paper is organized as follows. First, the surface model is presented in Sec. 2. Section 3 presents a case study to compare the proposed modeling approach with existing methods. Then, Sec. 4 discusses the selection of hyperparameters, effects of the sample size and the number of tasks, and the effect of the correlation strength between the machining process variable and surface height. Finally, Sec. 5 concludes the paper.

2 Surface Shape Modeling by Integrating Cutting Force Variation With Multitask Learning

In this section, the engineering-guided multitask learning (EG-MTL) surface model which integrates engineering physics with multitask learning is first presented, and then, an iterative algorithm is developed to estimate the parameters of the surface model.

2.1 Engineering-Guided Multitask Learning Surface Model. Traditional multitask learning approaches for Gaussian processes are data-driven, and do not incorporate engineering physics into the model. This paper develops a new EG-MTL surface model that integrates cutting force variation modeling with multitask learning, aiming to improve the surface prediction accuracy based on limited measurement data. Assume there are m similar-but-not-identical surfaces or surface manufacturing processes, and for surface l , the surface model takes the following form in the equation given below:

$$Z_l(\mathbf{s}) = f(\mathbf{U}_l(\mathbf{s}); \boldsymbol{\beta}_l) + \eta_l(\mathbf{s}) + \varepsilon \quad (1)$$

where l is the surface index, and $l = 1, \dots, m$; $\mathbf{s} = (x, y)$ is the coordinate vector, $Z_l(\mathbf{s})$ is the surface height at location \mathbf{s} in surface l , $\mathbf{U}_l(\mathbf{s})$ is a covariate vector of the location-dependent highly correlated process variables identified by expert knowledge, $f(\cdot)$ is a generic function capturing the correlation between the global trend of the surface and process specific engineering factors, e.g., material removal rate, $\eta_l(\mathbf{s})$ is a zero-mean Gaussian process, and $\varepsilon \sim \mathcal{N}(0, \psi^2)$.

In model (1), the surface variation is decomposed into a global trend that is induced by process settings and a local variation part that is modeled by a zero-mean Gaussian process. The global trend is modeled by a linear combination of multiple process variables. When there are both categorical and continuous process variables, a generalized linear model may be used instead [22,23]. The local variation is modeled as a zero-mean Gaussian process, and will be jointly learned across multiple similar-but-not-identical surfaces.

The rationale behind this decomposition is as follows. First, model $f(\mathbf{U}_l(\mathbf{s}); \boldsymbol{\beta}_l)$ is employed to capture the relationship between the process inputs and the surface height in face milling processes. It is reported that the surface height variation in a face milling process is strongly influenced by a number of engineering factors [13,20,21], including (1) the product/surface design that characterizes the design patterns of a surface, e.g., size, shape, and spatial distribution of holes and slots, (2) physical attributes of part materials, such as the defects and heterogeneous physical attributes caused by manufacturing flaws from suppliers, (3) manufacturing process conditions, such as feed rate, spindle tilt, spindle speed, depth of cut, cutter path, and clamping force, and (4) multistage interdependence that characterizes the effects of downstream stages on the surface shapes created in the upstream stages. For details, please refer to Ref. [13]. The functional form of $f(\cdot)$ should be determined based on a thorough understanding of the process physics. For instance, in a face milling process, the global

trend can be approximated by a linear term [20,13], i.e., $f(\mathbf{U}(\mathbf{s}); \boldsymbol{\beta}) = [\mathbf{1}, \text{MRR}(\mathbf{s})]\boldsymbol{\beta}$, where $\mathbf{1}$ is a column vector of ones, $\text{MRR}(\mathbf{s})$ is a column vector and represents the material removal rate (MRR) at location \mathbf{s} , and $\boldsymbol{\beta} = [\beta_0, \beta_1]^T$ is a 2×1 vector. This linear representation reveals the fact that higher axial cutting force variations due to MMR variation may result in large displacement between the surface and the cutter, causing significant surface height variations. Second, Gaussian process has been proven to be an effective method to model the local random surface variation that exhibits the spatial distribution similarity [13,24,25].

Note: The task similarity/relatedness should be carefully evaluated prior to applying model (1). However, the definition of task relatedness is unclear [26]. In the applications where expert knowledge is available, task relatedness may be defined based on the feedback of experts [26,27]. In machining processes, model (1) may be applied to surfaces machined using the same machine or parallel machines in a manufacturing plant. When expert knowledge is limited or unavailable, a preliminary study can be conducted to evaluate the benefits brought by using MTL.

2.2 Iterative Algorithm for Parameter Estimation of the EG-MTL Model. The parameters in model (1) that need to be estimated include: (1) the parameter vector in the global trend, $\boldsymbol{\beta}_l$, and (2) the parameters in the zero-mean Gaussian process, $\eta_l(\mathbf{s})$, where $l = 1, \dots, m$. The global trend describes the overall spatial variation pattern on a surface, and is mostly determined by the process settings. Thus, $\boldsymbol{\beta}_l$ reflects the overall surface variation and is not location dependent, but it varies with different surfaces. The local variation part shares similarity across multiple surfaces, and multitask learning can potentially improve the estimation accuracy. Figure 4 illustrates the scheme for learning the parameters. As indicated by Fig. 4, three information sources are fused in the spatial model, namely, (1) surface measurement from the process of interest, (2) cutting force variation, and (3) measurement data from other similar-but-not-identical processes. A summary of the difference between the proposed model and existing approaches is given below.

- Traditional multitask learning approaches only employ information from Eqs. (1) and (3), and do not incorporate surface cutting physics;
- Simple kriging interpolation adopts data source (1) only; and
- More advanced kriging approaches, e.g., co-kriging and kriging with external drift (KED), jointly utilize information sources (1) and (2), but cannot transfer knowledge from other relevant processes.

In the remainder of this section, we first review the multitask learning algorithm for the joint learning of multiple zero-mean Gaussian processes, and then, present an iterative algorithm for estimating the parameters in model (1).

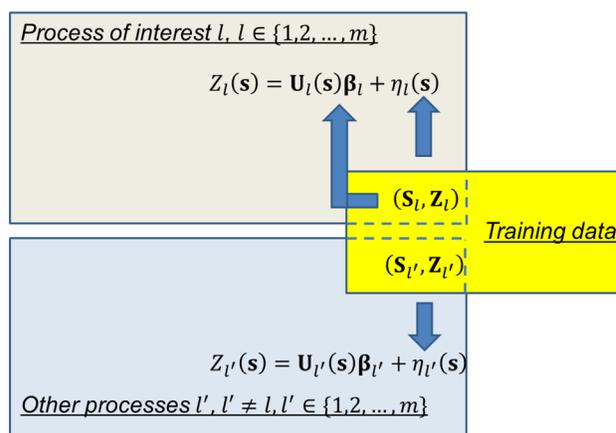


Fig. 4 Illustration of the EG-MTL scheme

2.2.1 Review of Multitask Learning. A number of multitask learning algorithms are available for Gaussian processes in the literature, such as Refs. [18,17]. The performance of different approaches depends on the data structure. Without losing generality, this study reviews the approach developed by Yu et al. [18] to illustrate the concept.

The goal of multitask learning for Gaussian processes is to estimate m related functions η_l , $l = 1, \dots, m$, based on training data represented by the following equation:

$$\mathbf{D}_l = (\mathbf{S}_l, \boldsymbol{\eta}_l) \quad (2)$$

where $\mathbf{S}_l \in \mathbb{R}^{n_l \times d}$ is a set of the inputs for function l , $\boldsymbol{\eta}_l \in \mathbb{R}^{n_l}$ is a set of the function values at these inputs, and n_l is the size of training data for task l . In surface modeling, the location is generally determined by two coordinates (x and y), and thus $d=2$. It is further assumed that there are in total n distinct data points in $\{\mathbf{D}_l\}$ with $\min(\{n_l\}) \leq n \leq \sum_{l=1}^m n_l$.

The commonality among different tasks is defined via the following inductive model [18].

EG-MTL Model: Let $\boldsymbol{\eta}^l$ be the values of η_l on a set \mathbf{S} , satisfying $\cup \mathbf{S}_l \subset \mathbf{S}$. Given the hyperprior distribution of $\boldsymbol{\mu}_z$ and \mathbf{C}_z described by a normal-inverse-Wishart distribution as shown by the following equation, the Gaussian process for task l is generated by the following three steps:

$$p(\boldsymbol{\mu}_z, \mathbf{C}_z) = \mathcal{N}\left(\boldsymbol{\mu}_z \mid 0, \frac{1}{\pi} \mathbf{C}_z\right) \mathcal{IW}\left(\mathbf{C}_z \mid \tau, \boldsymbol{\kappa}^{-1}\right) \quad (3)$$

- (1) $\boldsymbol{\mu}_z, \mathbf{C}_z$ are generated once using Eq. (3)
- (2) For each function η_l

$$\boldsymbol{\alpha}^l \sim \mathcal{N}(\boldsymbol{\mu}_z, \mathbf{C}_z) \quad (4)$$

- (3) Given $\mathbf{s} \in \mathbf{S}_l$

$$\eta_l(\mathbf{s}) = \sum_{i=1}^n \alpha_i^l \kappa(\mathbf{s}_i, \mathbf{s}) + \varepsilon \quad (5)$$

where $\kappa(\cdot, \cdot)$ is the base kernel function; $\mathbf{s} \in \mathbf{S}$; and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. It should be noted that the model parameters follow different distributions (Eq. (4)) in different tasks, and the parameters of the distributions follow a common hyper prior distribution (Eq. (3)). The hyper prior distribution reflects the commonality of tasks.

The estimates of $\Theta = \{\boldsymbol{\mu}_z, \mathbf{C}_z, \sigma^2\}$ can be obtained using an expectation-maximization (EM) algorithm [18], and the details of the EM procedure are presented in the Appendix. After obtaining Θ and $\{\hat{\boldsymbol{\alpha}}^l\}$ with the EM algorithm, the function value corresponding to an unknown input \mathbf{s}_u from task l is estimated using the following equation:

$$\hat{\eta}_l(\mathbf{s}_u) = \sum_{i=1}^n \hat{\alpha}_i^l \kappa(\mathbf{s}_i, \mathbf{s}_u) \quad (6)$$

where the prediction is calculated as a linear combination of $\kappa(\mathbf{s}_i, \mathbf{s}_u)$, and $\mathbf{s}_i \in \mathbf{S}$ includes both the training data from task l and these from other tasks. The kernel function $\kappa(\cdot, \cdot)$ is a measure of the “distance” or similarity between two inputs.

One critical limitation of this approach is that the surface means must be accurately estimated and removed prior to conducting multitask learning, because it is based on the assumption that the Gaussian processes have zero means. However, estimation of the global means with limited training data is often not sufficiently accurate. The prediction results can be misleading when the zero-mean assumption does not hold. This paper utilizes a linear model to characterize the global trend with expert process knowledge. The estimation is expected to be more accurate, and the zero-mean assumption for residual local variation will be valid.

2.2.2 Iterative Multitask Learning Algorithm. There are two types of model parameters in model (1), i.e., coefficients in the linear model, $\boldsymbol{\beta}_l$, and Gaussian process parameters. Simultaneously estimating both types of parameters is challenging as the changes in one part directly impacts the estimation of the other. To address this issue, an iterative algorithm is developed for the model estimation. The algorithm is illustrated by Fig. 5. For m similar-but-not-identical surfaces, the global trend and the Gaussian processes are estimated in an iterative manner, and this process is terminated when the coefficients in the global trend converge. In this procedure, the superscript j ($j=0, 1, \dots$) specifies the iteration index, and the subscript l ($l=1, \dots, m$) indicates the task/part/surface number. The procedure is further explained as follows.

As an initialization action, the Gaussian process part and the coefficient vector in the global trend are both set to be zero for all surfaces, as shown by Eq. (7). In general, selection of the initial values do not affect the prediction performance of the algorithm but will affect the convergence speed.

$$\begin{aligned} \hat{\eta}_l^0(\mathbf{s}) &= 0 \\ \hat{\boldsymbol{\beta}}_l^0 &= 0 \end{aligned} \quad (7)$$

where $l = 1, \dots, m$.

In the j th iteration, the Gaussian process part estimated from $(j-1)$ th iteration, $\hat{\eta}_l^{j-1}(\mathbf{s})$, is first removed from the surface height for all surfaces, as shown by Eq. (8). Then, robust linear regression is applied to estimate the coefficient vector $\hat{\boldsymbol{\beta}}_l^j$ for $l = 1, \dots, m$. The model of the global trend for surface l at the j th iteration is given by Eq. (9).

$$\hat{\mu}_l^j(\mathbf{s}) = Z_l(\mathbf{s}) - \hat{\eta}_l^{j-1}(\mathbf{s}) \quad (8)$$

where $Z_l(\mathbf{s})$ is the height of the l th surface; $\hat{\eta}_l^{j-1}(\mathbf{s})$ is the estimated Gaussian process for l th surface at the $(j-1)$ th iteration; and $l = 1, \dots, m$

$$\hat{\mu}_l^j(\mathbf{s}) = \mathbf{U}_l(\mathbf{s}) \hat{\boldsymbol{\beta}}_l^j \quad (9)$$

where $l = 1, \dots, m$.

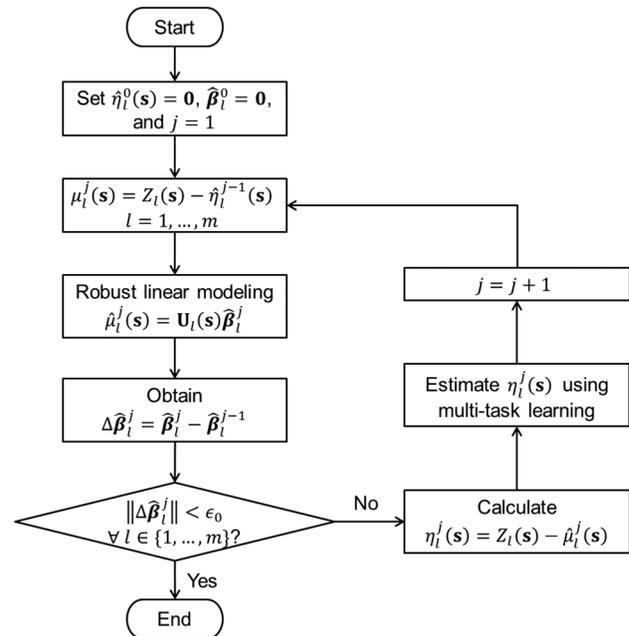


Fig. 5 Flowchart for the iterative algorithm

There are mainly three categories of robust regression methods for linear models: (1) least squares alternatives [28], (2) parametric alternatives [29], and (3) unit weights [30]. This study adopts the method presented in Ref. [28], which applies iteratively reweighted least squares for robust regression. This method has been implemented in some popular computing softwares, such as R and MATLAB. [31] provided a comprehensive overview of the popular robust methods.

After the robust linear regression, the convergence test is conducted to check whether the coefficients of the linear models converge. The change of the coefficient vector is defined by Eq. (10). The convergence criterion is given by Eq. (11). The convergence is considered to be achieved if the changes of the coefficient vectors for all surfaces are within a predetermined threshold

$$\Delta \hat{\beta}_l^j = \hat{\beta}_l^j - \hat{\beta}_l^{j-1} \quad (10)$$

where $\hat{\beta}_l^j$ and $\hat{\beta}_l^{j-1}$ are the coefficient vectors from iterations j and $j-1$, respectively; and $l = 1, \dots, m$.

$$\|\Delta \hat{\beta}_l^j\| < \varepsilon_0, \quad \forall l \in \{1, \dots, m\} \quad (11)$$

where ε_0 is a predetermined threshold and can be tuned based on applications and the accuracy requirement.

If convergence is achieved at the j th iteration, the parameter estimation is finished, and the prediction results at the $(j-1)$ th iteration are the final results of the algorithm. If convergence is not achieved, multitask Gaussian process learning (MTGPL) is then conducted to jointly estimate $\eta(\mathbf{s})$'s.

The detailed implementation of multitask learning for Gaussian processes is illustrated by Fig. 6. Specifically, the first step is to assign the hyper-prior parameters, including τ and π in the Normal-inverse-Wishart distribution as described by Eq. (3) and to determine the base kernel function form $\kappa(\cdot, \cdot)$. A common practice is to apply cross-validation to select the best settings from a set of candidate parameters and function forms [18,32]. The second step is to estimate the model and perform predictions, and this step follows the procedure reviewed in Sec. 2.2.1.

The computational complexity of the EG-MTL model is $O(k_1 k_2 m n^3)$, where k_1 is the number of iterations for EG-MTL parameter estimation, k_2 is the number of EM iterations for MTL, m is the number of tasks, and n is the training sample size. Here, the computational complexity of robust linear regression is neglected, because the MTL estimation is more dominant. The proposed method is less efficient than Gaussian process multitask learning (GPMTL), the complexity of which is $O(k_2 m n^3)$ [18], and simple kriging. But it needs to be pointed out that GPMTL and simple kriging generally yield larger prediction errors than

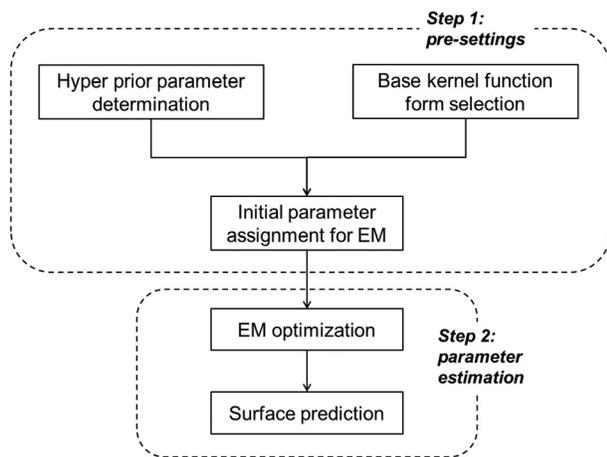


Fig. 6 Implementation procedure of multitask learning

the EG-MTL model, which will be shown by the case study in Sec. 3 and the discussion in Sec. 4. To overcome the computational challenge, we will discuss the economic sample size in Sec. 4.2. In practice, a tradeoff between the computational efficiency and prediction performance can be obtained by conducting a prior study that is similar with that in Sec. 4.2.

It should be noted that inappropriate initial settings may lead to undesirable results, e.g., slow convergence or nonconvergence and unsatisfactory prediction accuracy. Several practical suggestions are provided as follows to improve the EM efficiency and guarantee the acquisition of the global maximum:

- random restart (starting with different randomly selected initial parameter values)
- simulated annealing [33,34]
- monitoring the log-likelihood trace, and changing the initial values when (1) fluctuation occurs or (2) convergence is not achieved within a certain number of iterations

3 Case Study

This section compares the proposed engineering-guided multitask learning surface model (denoted as “EG-MTL model” afterwards) with some representative interpolation methods which have been applied popularly, including the multitask learning for Gaussian processes (referred to as GPMTL), simple kriging, and kriging with external drift (KED). A summary of these methods is given in Table 1.

GPMTL applies the algorithm described in Sec. 2.2.1. However, this algorithm is only applicable to zero-mean Gaussian processes [18]. The global mean is estimated using the sample mean from the training data. For surface $l, l = 1, \dots, m$, the global mean is estimated using the following equation:

$$\hat{\mu}_{l,0} = \bar{Z}(\mathbf{s}_{l,o}) = \frac{\sum_{\mathbf{s} \in \mathbf{S}_l} Z(\mathbf{s})}{n_l} \quad (12)$$

where $\mathbf{s}_{l,o}$ represents the observed locations in surface l . After removing the global mean, the residuals $\eta(\mathbf{s})$'s are jointly estimated using the MTGPL algorithm in Sec. 2.2.1.

KED is a generalized case of kriging where the overall trend is modeled as a function of some auxiliary predictors [35,36]. Linear models may be applied to capture the linear correlations between the target variable and the auxiliary predictors [36]. In the interpolation problem discussed in this study, a linear model is proposed since there is a strong correlation between the surface height Z and the process variables \mathbf{U} . This interpolation technique incorporates cutting force variation but do not employ the similarity between the process of interest and other similar-but-nonidentical processes.

A laser holographic interferometer was adopted at a Ford engine plant to measure engine block deck surfaces, and the obtained surface data from this plant will be used throughout this section. The root mean squared error (RMSE) will be used as a metric of the prediction performance, and its resolution is given by Eq. (13). Another metric of interest is the improvement

Table 1 Method summary for the case study

Method	Using cutting force variation modeling?	Transferring knowledge from other processes?
EG-MTL	Yes	Yes
GPMTL	No	Yes
Simple kriging	No	No
KED	Yes	No

Table 2 Summary of part information

Part number (l)	1	2	3
Training sample size (n_l)	50	50	50
Test sample size ($ \mathcal{G}_l $)	32563	32790	32673
Correlation strength ($\rho(U, Z)$)	0.84	0.78	0.78

percentage of RMSE from the simple kriging method to the EG-MTL model or MTGPL, which is defined by Eq. (14)

$$\text{RMSE}(l) = \sqrt{\frac{\sum_{s \in \mathcal{S}} (\hat{Z}_l(s) - Z_l(s))^2}{|\mathcal{G}_l|}} \quad (13)$$

where $l = 1, \dots, m$ is the task/surface number, \mathcal{G}_l is a set of the locations where prediction is made, and $|\mathcal{G}_l|$ denotes the size of \mathcal{G}_l , i.e., test sample size for surface l .

$$\Delta_+ \text{RMSE}_{\text{kriging}} = \frac{\text{RMSE}_{\text{kriging}} - \text{RMSE}_{\text{EG-MTL}}}{\text{RMSE}_{\text{kriging}}} \times 100\% \quad (14)$$

where $\Delta_+ \text{RMSE}_{\text{kriging}}$ is the improvement percentage of RMSE from simple kriging to the EG-MTL model; and $\text{RMSE}_{\text{kriging}}$ and $\text{RMSE}_{\text{EG-MTL}}$ are the RMSEs of simple kriging and the proposed EG-MTL model, respectively.

Similarly, a metric is defined for the RMSE improvement from the KED interpolation to the EG-MTL model, which can be viewed as a quantification of the benefits brought by multitask learning when expert knowledge is adopted. The metric is given by the following equation:

$$\Delta_+ \text{RMSE}_{\text{KED}} = \frac{\text{RMSE}_{\text{KED}} - \text{RMSE}_{\text{EG-MTL}}}{\text{RMSE}_{\text{KED}}} \times 100\% \quad (15)$$

where $\Delta_+ \text{RMSE}_{\text{KED}}$ is the improvement percentage of RMSE from KED to the EG-MTL model; and RMSE_{KED} and $\text{RMSE}_{\text{EG-MTL}}$ are the prediction RMSEs of KED and the EG-MTL model, respectively.

The measurement data of three engine head surfaces are utilized to validate the proposed modeling approach. The part information is summarized by Table 2. The test data of each part consists of approximately 32,600 points, and 50 points are randomly sampled as training data. Here, the random sampling aims to generate data mimicking measurements from a low-resolution metrology system. It is reasonable to sample the data that characterizes major variation features based on process knowledge (e.g., the surface height variation between cylinder bores). Our method

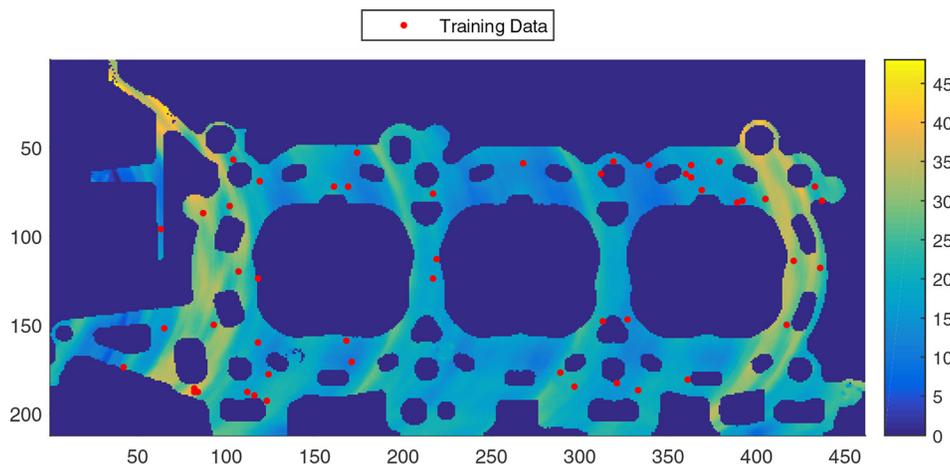


Fig. 7 An engine head surface example

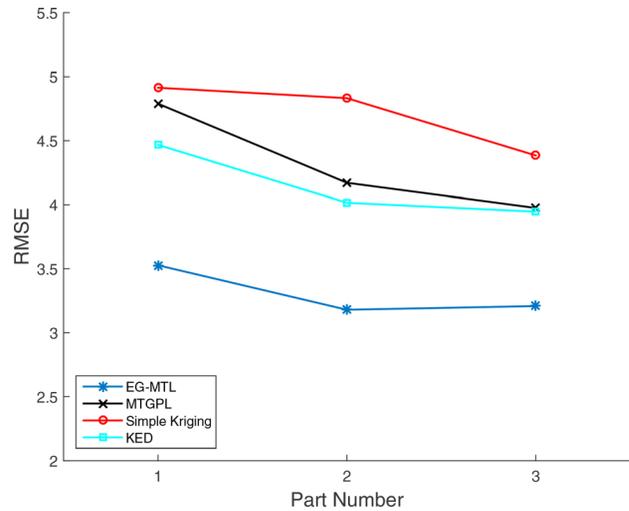


Fig. 8 RMSE comparison for the EG-MTL model, GPMTL, and kriging

is then applied to the downsampled data, i.e., training data, for prediction and validation. Figure 7 shows an engine head surface example, where the height is represented by color, namely, dark-scale shading indicates small value and bright-scale shading corresponds to large value; and dots represent sampling locations. The machined surfaces all exhibit a high-low-high-low trend along the horizontal direction; however, these patterns are not identical. For each part, a covariate, U , is simulated to have a correlation of around 0.80 with the surface height, Z . Additionally, in this case study and Sec. 4, the unit of the measurement data and the prediction error is μm . We will omit the unit from now on for the sake of simplicity. The value of ϵ_0 in the convergence criterion (Eq. (11)) is chosen to be 0.1 and remain the same throughout the case study.

The comparative results are shown by Fig. 8. The average RMSEs of the EG-MTL model, GPMTL, simple kriging, and KED over three parts are 3.30, 4.31, 4.71, and 4.14, respectively. It is seen that for all three parts, the EG-MTL model performs the best, and the simple kriging obtains the worst prediction accuracy. The KED model yields smaller RMSEs than MTGPL and simple kriging, and this result implies that the incorporation of cutting force variation modeling is beneficial. Hence, by incorporating a highly correlated covariate, which can be identified using engineering understanding of the process physics, the EG-MTL model is able to significantly improve the prediction accuracy over existing interpolation techniques. When such knowledge is not available, the data-driven model

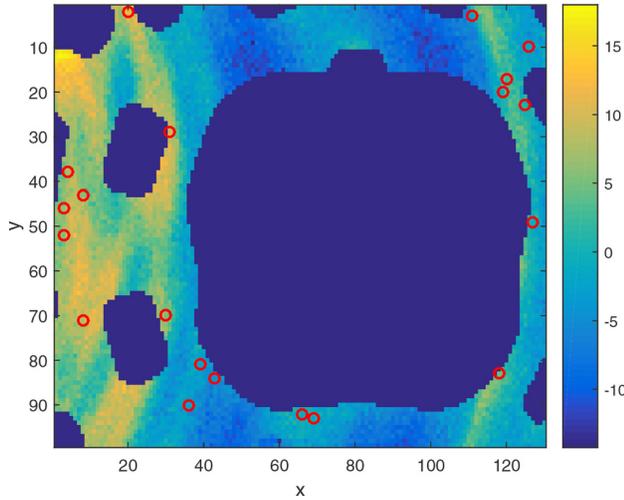


Fig. 9 Engine surface example for the hyperparameter study

Table 3 Experimental setting for the hyperprior parameter study

Number of tasks (m)	5
Training sample size (n_t)	20
Test sample size ($ \mathcal{G}_t $)	Approximately 11000
Candidate hyperparameter values	0.0001, 0.001, 0.01, 0.1, 0.5, 1, 6, ..., 151
Total number of experimental runs	46656

reviewed in Sec. 2.2.1 can be used, because GPMTL outperforms simple kriging, as indicated by Fig. 8.

4 Discussion

This section discusses the following three issues: (1) hyperparameter selection, (2) effects of the sample size and the number of

tasks, and (3) the effect of the correlation strength between the covariate and the surface height.

4.1 Hyperparameter Selection. It was found by a preliminary study that a quadratic exponential kernel function performed well in modeling surface height, and we will use this kernel form in this study. Quadratic exponential kernel functions are popularly used in spatial statistics, and the general form is given by the following equation:

$$\kappa(\mathbf{s}, \mathbf{s}') = \exp\left(\frac{-\|\mathbf{s} - \mathbf{s}'\|^2}{\delta^2}\right) \quad (16)$$

where δ^2 is an unknown parameter and will be determined later.

Three hyperprior parameters need to be determined prior to the modeling, i.e., τ and π from the normal-inverse-Wishart distribution, as well as δ^2 from the base kernel function. An experiment is conducted to investigate the effects of the hyper parameters using a part of the engine surface. Figure 9 displays an example of the surface, where circles indicate sampling location. A summary of the experimental setting is given as Table 3. In this study, the number of tasks and the training sample size are fixed as 5 and 20, respectively; the test sample size is about 11,000; the candidate parameter values vary from 0.0001 to 151 (in total 36 values); and all possible combinations (in total $36^3 = 46,656$) were tested.

The average RMSE for simple kriging is 6.91, and the best parameter combination found for multitask learning is $[\tau, \pi, \delta^2] = [0.0001, 1, 111]$ with a prediction error of 2.87, which is much smaller than that of simple kriging. Figure 10 shows the 1D RMSE trends versus τ , π , and δ^2 . Figure 11 displays the RMSE trends when changing the pairs of $[\tau, \pi]$, $[\tau, \delta^2]$, and $[\pi, \delta^2]$.

Figure 10(a) shows that as τ increases, the RMSE first decreases sharply and then increases slowly, meaning that the prediction performance first improves significantly and then degrades. A similar pattern is observed for π , as displayed by Fig. 10(b). From Fig. 10(c), it is noted that when increasing δ^2 , the RMSE first decreases, and then increases slowly, meaning that

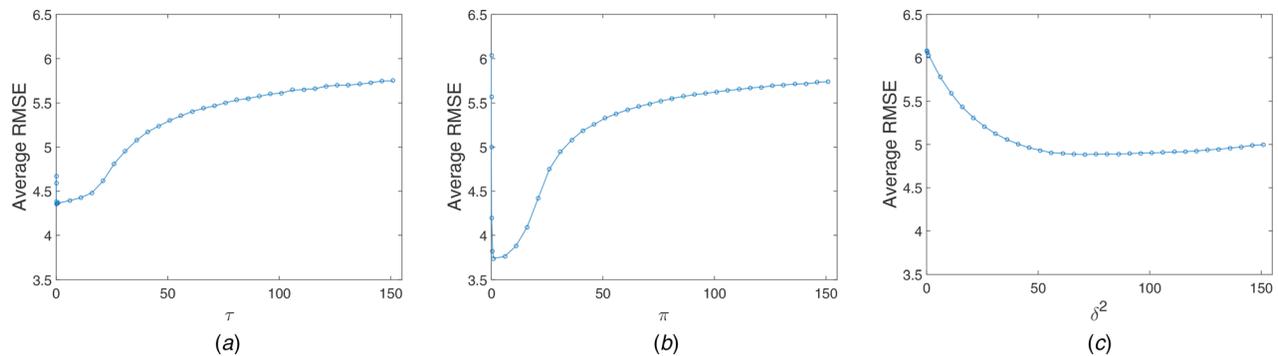


Fig. 10 Effects of hyperparameters on RMSE: (a) RMSE versus τ , (b) RMSE versus π , and (c) RMSE versus δ^2

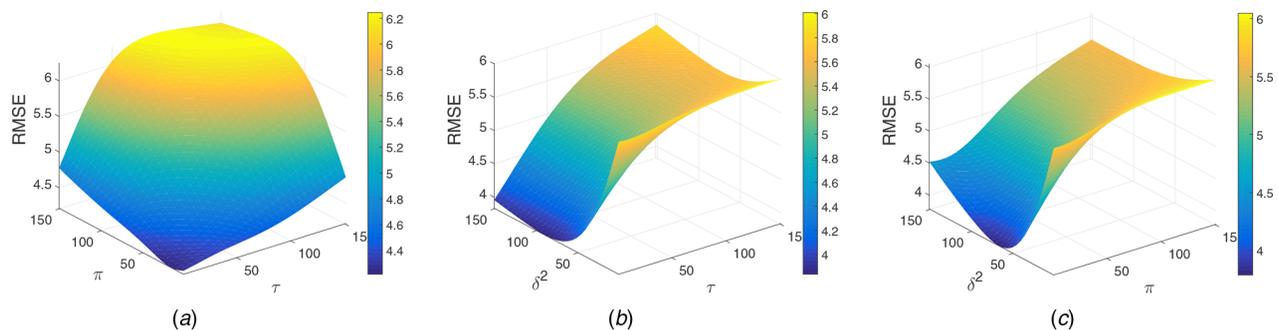


Fig. 11 Effects of hyperparameter pairs on RMSE: (a) RMSE versus (τ, π) , (b) RMSE versus (τ, δ^2) , and (c) RMSE versus (π, δ^2)

Table 4 Summary of experimental settings for the effects of the sample size and the number of tasks

Experiment	1	2	3
Number of tasks (m)	5	3, 4, ..., 20	3, 4, ..., 10
Training sample size (n_i)	10, 50, 100, ..., 300	50	10, 20, ..., 150
Test sample size ($ \mathcal{G}_t $)	Approximately 32,600	Approximately 32,600	Approximately 32,600
Number of repeated runs	10	10	1

the prediction performance does not change significantly when δ^2 is at a relatively high level, e.g., $\delta^2 > 50$. These patterns can be reconfirmed by the 3D RMSE plots which are shown by Figs. 11(a)–11(c). It is concluded that

- (1) In general, smaller τ and π along with larger δ^2 yield better prediction performance.
- (2) In practice, it is suggested to choose τ and π to be around 1, and δ^2 to be between 50 and 100.

The selection of optimal hyperparameter is mostly determined by the spatial variation pattern. When implementing the proposed modeling approach in practice, a cross-validation study may be conducted using the training data to find appropriate values.

4.2 Effects of the Sample Size and the Number of Tasks.

The learning performance of multitask learning can be sensitive to the sample size and the number of tasks. Therefore, it is important to analyze the effects of these factors in practice. In this section,

three experiments are conducted to investigate the effects of the sample size and the number of tasks. A summary of the experimental settings is given by Table 4.

In the first experiment, the number of tasks is fixed at 5, and the sample size varies from 10 to 300. Ten runs are repeated for every sample size, and the sampling locations are randomly selected for each run. The experiment results are illustrated by Fig. 12. It is seen that as the sample size increases, the prediction performance of multitask learning and kriging both improve, and Δ_+ RMSE first increases, then keeps stable, and shows a sign of decreasing in the end. Further, increasing the sample size may not be economically necessary and potentially cause overfitting issue. It is also noticed that the sampling of data will influence the prediction performance. When the experiments were repeated, the training data were randomly generated each time, and the performance variation is reflected by the boxplots in Fig. 12.

The second experiment fixes the sample size as 50 for all parts, and varies the number of tasks from 3 to 10. The results are shown

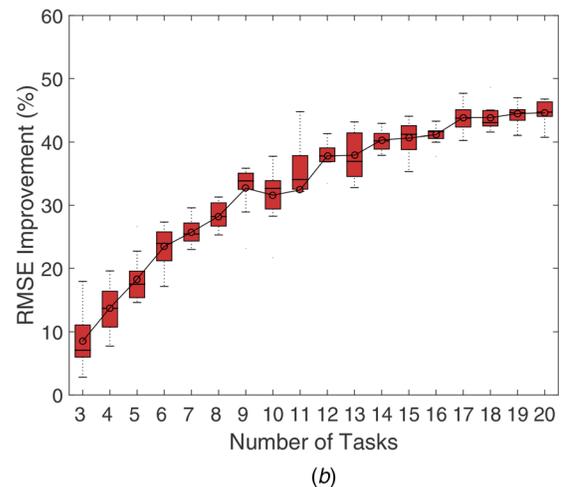
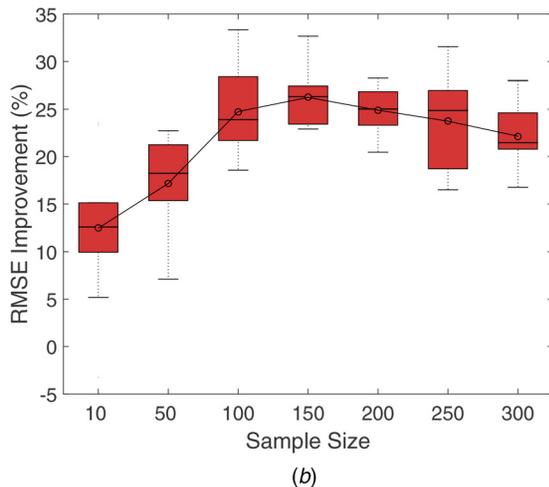
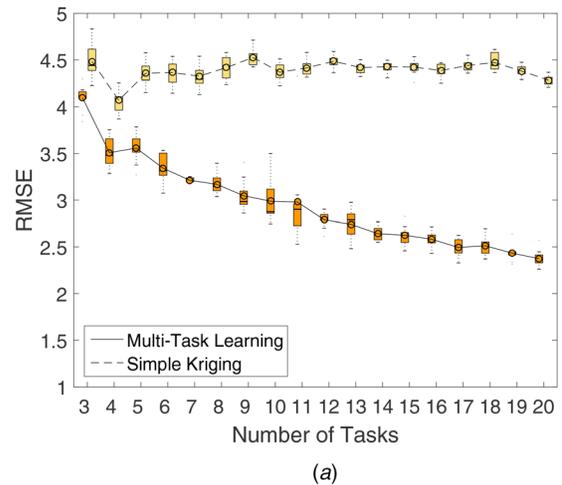
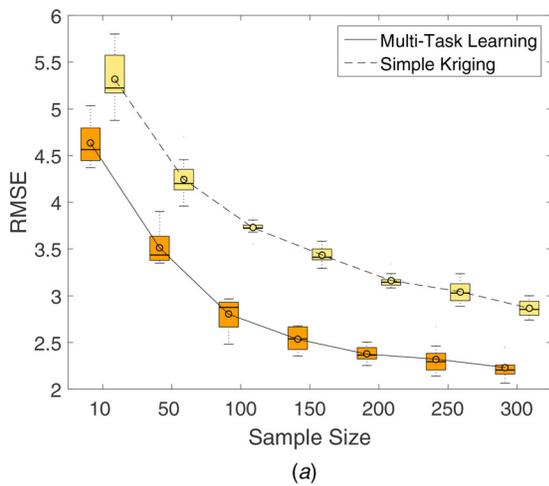


Fig. 12 Effects of sample size on the prediction performance: (a) RMSE and (b) Δ_+ RMSE

Fig. 13 Effects of number of tasks on the prediction performance: (a) RMSE and (b) Δ_+ RMSE

by Fig. 13. As the increase of the number of tasks, Δ_+RMSE first increases quickly, and then keeps stable after the number of tasks reaches 17.

In the third experiment, both the number of tasks and the sample size are changed. The number of tasks takes the values of {3, 4, ..., 10}, and the sample size varies as {10, 20, ..., 150}. Figure 14(a) displays the 3D trend of RMSE when varying the sample size and the number of tasks. Clearly, the RMSE decreases with greater sample size and number of tasks. Figure 14(b) shows the 3D plot of Δ_+RMSE , and it is indicated that the multitask learning outperforms simple kriging in all cases, and Δ_+RMSE increases as the increase of sample size and number of tasks.

The above experimental results lead to the following conclusions:

- (1) Multitask learning is able to improve the prediction accuracy over simple kriging;
- (2) The prediction accuracy improves as the increase of the training sample size for both multitask learning and kriging;
- (3) The prediction performance improves when the number of tasks increases for multitask learning;
- (4) Δ_+RMSE becomes bounded when the sample size and the number of tasks are relatively large.

Intuitively, when the training sample size is sufficient, single-task learning, which is simple kriging in this application, will be able to learn the spatial model very well, and the benefits brought by transferring knowledge thus become limited. Moreover, when the number of tasks increases, duplicate information may exist across tasks, and the improvement becomes less significant.

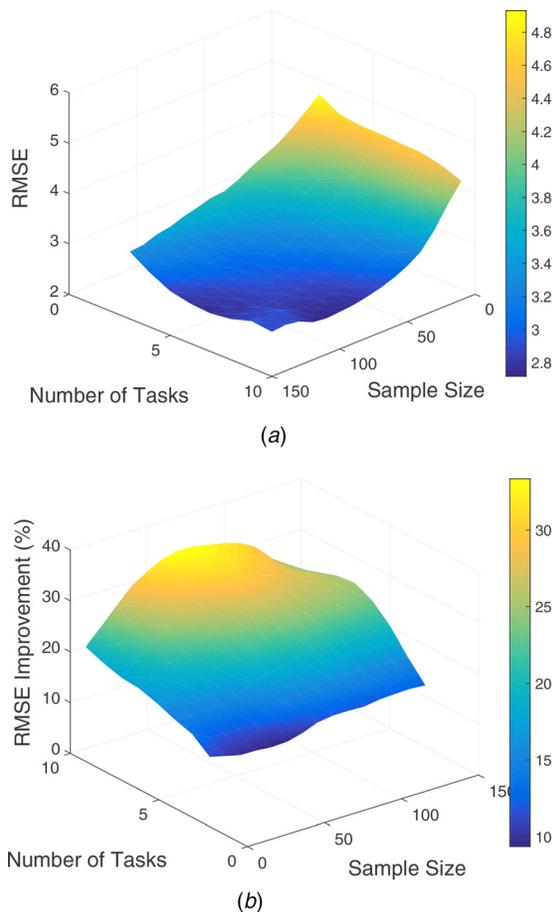


Fig. 14 Effects of sample size and number of tasks on the prediction performance: (a) RMSE and (b) Δ_+RMSE

4.3 Effect of the Correlation Strength. To investigate the influence of $\rho(U, Z)$ on the estimation accuracy, we vary the strength of $\rho(U, Z)$, and compare the performance of the EG-MTL model, MTGPL, simple kriging, and the KED method. The comparative results are illustrated using Figs. 15(a) and 15(b), which show the trends of RMSE and Δ_+RMSE when changing the correlation strength, respectively. In Figs. 15(a) and 15(b), the results corresponding to each correlation strength are averaged over three parts. The following observations are made from these results.

- (1) The performance of KED is strongly influenced by the correlation strength. When the correlation between U and Z is not sufficiently strong, e.g., $\rho(U, Z) < 0.75$, its prediction accuracy is worse than simple kriging. When $\rho(U, Z)$ becomes stronger, the prediction accuracy of KED improves, and outperforms simple kriging. Also, the RMSE improvement from KED to the EG-MTL model decreases as $\rho(U, Z)$ increases, indicating that the benefits introduced by multitask learning become more limited.
- (2) The prediction accuracy using the EG-MTL model is significantly affected by the correlation strength, i.e., as $\rho(U, Z)$ increases, RMSE for the EG-MTL model decreases

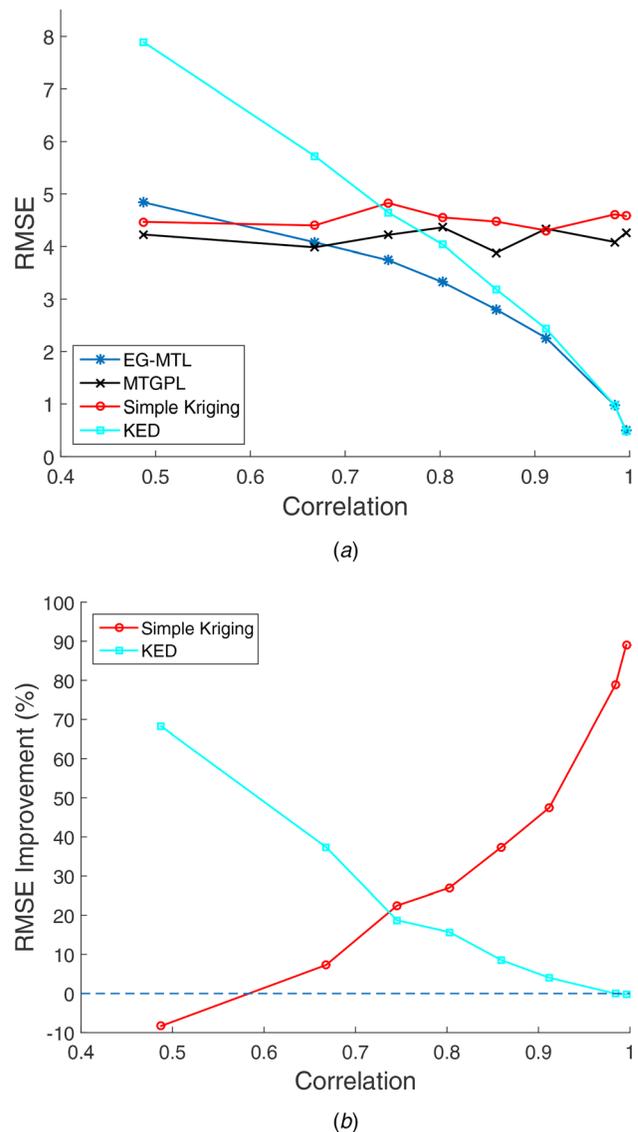


Fig. 15 Effects of $\rho(U, Z)$ on the prediction performance: (a) RMSE versus $\rho(U, Z)$ and (b) Δ_+RMSE versus $\rho(U, Z)$

and Δ_+ RMSE increases. When the correlation is relatively weak, e.g., less than 0.5, the introduction of such a covariate will degrade the modeling performance.

- (3) The proposed engineering guided GPMTL outperforms simple kriging in all cases, which means that when expert knowledge on the process is not available and highly correlated variables cannot be identified, the adoption of multi-task learning is still beneficial compared with single task learning.

5 Conclusion

The accomplishments of this paper can be summarized as follows:

- (1) An improved surface shape modeling approach by incorporating engineering insights with multitask learning for Gaussian processes. The uniqueness of the developed model is that information on both the cutting force variation modeling and spatial data from other similar-but-not-identical processes is fused in one framework. The model is advantageous in predicting surface height at unobserved locations over the traditional approaches, such as simple kriging, kriging with external drift, and multitask learning for zero-mean Gaussian processes. The approach is validated using the data from real-world machining processes in an engine plant.
- (2) An iterative algorithm for estimating model parameters. A novel algorithm is developed to iteratively estimate the parameters in the global trend and the local variation represented by zero-mean Gaussian processes. Parameters are updated in each iteration and the procedure is terminated once the convergence criterion is met. Moreover, practical guidelines are proposed to expedite the EM optimization process.
- (3) Systematic studies on the hyperparameter selection and the effects of training data size. Response surfaces of the prediction performance are created for the model hyperparameters, the number of tasks, and the training sample size.

The proposed engineering integration model is expected to improve the surface prediction accuracy when limited measurement data are available, thus enabling cost-effective fine-resolution surface characterization.

Future research will be focused on (1) variable selection for the iterative engineering guided GPMTL model and (2) task selection for multitask learning. First, the variables that reflect multidisciplinary process information may take various types and formats. These variables may not be equally important for improving the surface shape prediction or they may contain duplicate information. Thus, a variable selection algorithm is needed to identify the most relevant variables to be incorporated in the surface model. Second, a systematic approach is desirable to select the tasks that are used for multitask learning. Some algorithms have been presented in the literature to select tasks based on the quantification of the similarity among tasks. However, the selection may be more complicated in the context of the proposed integration model, and nontrivial efforts are needed to develop a systematic way for the task selection.

Acknowledgment

This research has been supported by an NSF GOALI Grant No. CMMI-1434411.

Appendix

The estimates of $\Theta = \{\mu_x, C_x, \sigma^2\}$ can be obtained using the following EM algorithm [18]. Detailed derivation of the algorithm is available in Ref. [18].

- *E-step*: Estimate the expectation and covariance of α , $l = 1, \dots, m$, given the current Θ .

$$\hat{\alpha}^l = \left(\frac{1}{\sigma^2} \kappa_l^T \kappa_l + C_x^{-1} \right)^{-1} \left(\frac{1}{\sigma^2} \kappa_l^T \eta_l + C_x^{-1} \mu_x \right) \quad (A1)$$

$$C_{x^l} = \left(\frac{1}{\sigma^2} \kappa_l^T \kappa_l + C_x^{-1} \right)^{-1} \quad (A2)$$

where $\kappa_l \in R^{m \times n}$ is the base kernel $\kappa(\cdot, \cdot)$ evaluated between S_l and S .

- *M-step*: Optimize $\Theta = \{\mu_x, C_x, \sigma^2\}$.

$$\mu_x = \frac{1}{\pi + m} \sum_{l=1}^m \hat{\alpha}^l \quad (A3)$$

$$C_x = \frac{1}{\tau + m} \times \left\{ \pi \mu_x \mu_x^T + \tau \kappa^{-1} + \sum_{l=1}^m C_{x^l} + \sum_{l=1}^m [\hat{\alpha}^l - \mu_x] [\hat{\alpha}^l - \mu_x]^T \right\} \quad (A4)$$

$$\sigma^2 = \frac{1}{m} \sum_{l=1}^m \sum_{n_l} \|\eta_l - \kappa \hat{\alpha}^l\|^2 + \text{tr}[\kappa_l C_{x^l} \kappa_l^T] \quad (A5)$$

where $\text{tr}(\cdot)$ is the trace operator.

References

- [1] Willomitzer, F., Ettl, S., Arold, O., and Häusler, G., 2013, "Flying Triangulation—A Motion—Robust Optical 3d Sensor for the Real-Time Shape Acquisition of Complex Objects," AIP Conference Proceeding, Vol. 1537, pp. 19–26.
- [2] Zhu, X., Ding, H., and Wang, M. Y., 2004, "Form Error Evaluation: An Iterative Reweighted Least Squares Algorithm," *ASME J. Manuf. Sci. Eng.*, **126**(3), pp. 535–541.
- [3] Yang, B.-D., and Menq, C.-H., 1993, "Compensation for Form Error of End-Milled Sculptured Surfaces Using Discrete Measurement Data," *Int. J. Mach. Tools Manuf.*, **33**(5), pp. 725–740.
- [4] Grove, D. M., Woods, D. C., and Lewis, S. M., 2004, "Multifactor b-Spline Mixed Models in Designed Experiments for the Engine Mapping Problem," *J. Qual. Technol.*, **36**(4), pp. 380–391.
- [5] D'Errico, J. R., 2005, "Surface Fitting Using Gridfit," The MathWorks, Inc., Natick, MA, <https://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit?requestedDomain=www.mathworks.com>
- [6] Yang, T.-H., and Jackman, J., 2000, "Form Error Estimation Using Spatial Statistics," *ASME J. Manuf. Sci. Eng.*, **122**(1), pp. 262–272.
- [7] Xia, H., Ding, Y., and Wang, J., 2008, "Gaussian Process Method for Form Error Assessment Using Coordinate Measurements," *IIE Trans.*, **40**(10), pp. 931–946.
- [8] Du, S., and Fei, L., 2016, "Co-Kriging Method for Form Error Estimation Incorporating Condition Variable Measurements," *ASME J. Manuf. Sci. Eng.*, **138**(4), p. 041003.
- [9] Bao, L., Wang, K., and Jin, R., 2014, "A Hierarchical Model for Characterising Spatial Wafer Variations," *Int. J. Prod. Res.*, **52**(6), pp. 1827–1842.
- [10] Plumlee, M., Jin, R., Roshan Joseph, V., and Shi, J., 2013, "Gaussian Process Modeling for Engineered Surfaces With Applications to SI Wafer Production," *Stat.*, **2**(1), pp. 159–170.
- [11] Wang, G., Gertner, G., and Anderson, A., 2005, "Sampling Design and Uncertainty Based on Spatial Variability of Spectral Variables for Mapping Vegetation Cover," *Int. J. Remote Sens.*, **26**(15), pp. 3255–3274.
- [12] Xiao, X., Gertner, G., Wang, G., and Anderson, A. B., 2005, "Optimal Sampling Scheme for Estimation Landscape Mapping of Vegetation Cover," *Landscape Ecol.*, **20**(4), pp. 375–387.
- [13] Suriano, S., Wang, H., Shao, C., Hu, S. J., and Sekhar, P., 2015, "Progressive Measurement and Monitoring for Multi-Resolution Data in Surface Manufacturing Considering Spatial and Cross Correlations," *IIE Trans.*, **47**(10), pp. 1033–1052.
- [14] Zhao, H., Jin, R., Wu, S., and Shi, J., 2011, "PDE-Constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process," *ASME J. Manuf. Sci. Eng.*, **133**(2), p. 021012.
- [15] Caruana, R., 1997, "Multitask Learning," *Mach. Learn.*, **28**(1), pp. 41–75.
- [16] Lawrence, N. D., and Platt, J. C., 2004, "Learning to Learn With the Informative Vector Machine," 21st International Conference on Machine Learning, ACM, p. 65.

- [17] Bonilla, E. V., Chai, K. M., and Williams, C., 2007, "Multi-Task Gaussian Process Prediction," *Advances in Neural Information Processing Systems*, pp. 153–160.
- [18] Yu, K., Tresp, V., and Schwaighofer, A., 2005, "Learning Gaussian Processes From Multiple Tasks," 22nd International Conference on Machine Learning, ACM, pp. 1012–1019.
- [19] Schwaighofer, A., Tresp, V., and Yu, K., 2004, "Learning Gaussian Process Kernels via Hierarchical Bayes," *Advances in Neural Information Processing Systems*, pp. 1209–1216.
- [20] Nguyen, H. T., Wang, H., and Hu, S. J., 2013, "Characterization of Cutting Force Induced Surface Shape Variation in Face Milling Using High-Definition Metrology," *ASME J. Manuf. Sci. Eng.*, **135**(4), p. 041014.
- [21] Nguyen, H. T., Wang, H., Tai, B. L., Ren, J., Hu, S. J., and Shih, A., 2016, "High-Definition Metrology Enabled Surface Variation Control by Cutting Load Balancing," *ASME J. Manuf. Sci. Eng.*, **138**(2), p. 021010.
- [22] Nelder, J. A., and Baker, R., 1972, "Generalized Linear Models," *Encyclopedia of Statistical Sciences*, Wiley, New York, NY.
- [23] Fahrmeir, L., Kneib, T., Lang, S., and Marx, B., 2013, "Generalized Linear Models," *Regression*, Springer, Berlin, pp. 269–324.
- [24] Stein, A., and Corsten, L., 1991, "Universal Kriging and Cokriging as a Regression Procedure," *Biometrics*, pp. 575–587.
- [25] Atkinson, P., Webster, R., and Curran, P., 1992, "Cokriging With Ground-Based Radiometry," *Remote Sensing of Environment*, **41**(1), pp. 45–60.
- [26] Chai, K. M., 2010, "Multi-Task Learning With Gaussian Processes," Ph.D. dissertation, University of Edinburgh, Edinburgh, Scotland.
- [27] Ghosh, J., and Bengio, Y., 2003, "Bias Learning, Knowledge Sharing," *IEEE Transactions on Neural Networks*, **14**(4), pp. 748–765.
- [28] Holland, P. W., and Welsch, R. E., 1977, "Robust Regression Using Iteratively Reweighted Least-Squares," *Commun. Stat. Theory Methods*, **6**(9), pp. 813–827.
- [29] Lange, K. L., Little, R. J., and Taylor, J. M., 1989, "Robust Statistical Modeling Using the t Distribution," *J. Am. Stat. Assoc.*, **84**(408), pp. 881–896.
- [30] Wainer, H., and Thissen, D., 1976, "Three Steps Towards Robust Regression," *Psychometrika*, **41**(1), pp. 9–34.
- [31] Maronna, R., Martin, D., and Yohai, V., 2006, *Robust Statistics*, Wiley, Chichester.
- [32] Huang, S., Li, J., Chen, K., Wu, T., Ye, J., Wu, X., and Yao, L., 2012, "A Transfer Learning Approach for Network Modeling," *IIE Trans.*, **44**(11), pp. 915–931.
- [33] Kirkpatrick, S., Gelatt, Jr., C. D., and Vecchi, M. P., 1983, "Optimization by Simulated Annealing," *Science*, **220**(4598), pp. 671–680.
- [34] Brooks, S. P., and Morgan, B. J., 1995, "Optimization Using Simulated Annealing," *J. R. Stat. Soc. (The Statistician)*, **44**(2), pp. 241–257.
- [35] Hudson, G., and Wackernagel, H., 1994, "Mapping Temperature Using Kriging With External Drift: Theory and an Example From Scotland," *Int. J. Climatol.*, **14**(1), pp. 77–91.
- [36] Bourennane, H., King, D., and Couturier, A., 2000, "Comparison of Kriging With External Drift and Simple Linear Regression for Predicting Soil Horizon Thickness With Different Sample Densities," *Geoderma*, **97**(3), pp. 255–271.